



(1)

Developing an Efficient Spectral Clustering Algorithm on Large Scale Graphs in Spark

Ahmed I. Taloba Marwan R. Riad Taysir Hassan A. Soliman

Abstract:

Recently, most of the data can be represented by graph structures, such as social media, Protein-Protein Interaction, transportation system, systems biology,...., etc. Many researches have been achieved to cluster very large graphs but more efficient algorithms are required since such a process takes a long time and requires more memory. In this paper, we propose an Efficient Spectral Clustering Algorithm on Large Scale Graphs in Spark (ESCALG), using map reduce function and shuffling phases in Dijkstra's algorithm. In addition, ESCALG depends mainly on a sparse matrix as a data structure, which less time in execution. Then, GraphX is applied to deal with graph data processing and in GraphX used Pregel in computing shortest path. To test the performance of ESCALG, it is compared with Large-Scale Spectral Clustering on Graphs and Standard Spectral Clustering Algorithms using seven datasets, where ESCALG proved high efficiency in terms of memory and time performance.

Keywords:

Spectral Clustering , Apache Spark, Large scale Graph Clustering

Published In:

2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS) , NULL , 292-298